# NETFLIX MOVIE DURATION ANALYSIS PROJECT REPORT

## Introduction

The Netflix Movie Duration Analysis project focuses on exploring and visualizing the trends in the average duration of movies available on the Netflix platform over the years. With Netflix becoming the largest entertainment/media company worldwide, it presents a unique opportunity to analyze the change in movie lengths and gain insights into the entertainment industry.

## Data Loading and Dictionary Creation

To begin the analysis, a Python dictionary was created to store the average movie durations for the years 2011 to 2020. The dictionary contains two lists, "years" and "durations," which store the corresponding data for each year.

## Creating a DataFrame from the Dictionary

The dictionary data was converted into a pandas DataFrame to facilitate further data manipulation and visualization. By using pandas, the movie durations and release years could be easily accessed and analyzed.

## Visual Inspection of Data

The DataFrame was visualized using a line plot to observe the trend in average movie durations over the specified years. This provided a clear indication that the average movie lengths have been decreasing during the period from 2011 to 2020.

## Loading Additional Data from CSV

To expand the analysis, the project acquired more data from a CSV file, containing a broader range of movie information. The new

DataFrame was created with columns such as "title," "country," "genre," "release_year," and "duration."

**Filtering for Movies**

To focus on movies only, the DataFrame was filtered to include only rows with "type" as "Movie." Subsequently, the columns of interest were selected, leading to the creation of a new DataFrame, "netflix_movies_col_subset."

**Creating a Scatter Plot**

Since the data now consists of individual movie entries, a scatter plot was used to visualize the duration of movies against their release years. The scatter plot provided a broader perspective on the distribution of movie durations over time.

**Digging Deeper**

To investigate the impact of non-typical genres on the declining average movie durations, the DataFrame was further filtered for movies with durations under 60 minutes. This led to the identification of genres such as "Children," "Stand-Up," and "Documentaries" that contributed to the observed trend.

**Marking Non-Feature Films**

To visually distinguish non-typical genres in the scatter plot, colors were assigned to different genres using a loop. This allowed for a more informative visualization and revealed the concentration of shorter films in certain genres.

**Conclusion**

The Netflix Movie Duration Analysis project successfully revealed a decline in the average duration of movies on the platform over the years. The exploration of genres also highlighted the influence of non-

typical genres, such as "Children," "Stand-Up," and "Documentaries," on the overall trend. Further analyses could be conducted to investigate the factors contributing to this trend and gain deeper insights into the movie landscape on Netflix.

Note: The project used Python libraries such as pandas and matplotlib.pyplot for data manipulation and visualization. The results obtained from the analysis are for exploratory purposes only and may require further investigation for comprehensive conclusions.

Thank you for taking the time to explore the Netflix Movie Duration Analysis project! If you have any questions or feedback, feel free to reach out to me at osuolalefolarin@gmail.com

GitHub Repository: https://github.com/Folarinosuolale/Data-Science-Machine-Learning